International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)

Vol. 10-No.4 (2024) pp. 1885-1892
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Heart Failure Prediction: A Comparative Study of SHAP, LIME, and ICE in Machine Learning Models

Tuğçe ÖZNACAR[1]*, Zeynep Tuğçe SERTKAYA[2]

[1]Ankara Medipol University, Faculty of Medicine, Department of Biostatistics, Ankara-Turkiye
* **Corresponding Author Email:** tugce.sencelikel@ankaramedipol.edu.tr - **ORCID:** 0000-0003-0364-0401

[2]Ankara Medipol University, Department of Physiology, Ankara-Turkiye
Email: zeynep.sertkaya@ankaramedipol.edu.tr - **ORCID:** 0000-0002-1307-6480

Article Info:

DOI: 10.22399/ijcesen.589
Received : 05 November 2024
Accepted : 29 December 2024

Keywords :

Machine Learning,
Heart Failure,
SHAP,
LIME,
ICE,

Abstract:

Heart failure remains a critical public health issue, prompting the need for effective predictive modeling. This study evaluates the performance of LightGBM, SVM, Random Forest, and Logistic Regression models on a heart failure dataset. Logistic Regression achieved the highest accuracy of 86.89%, demonstrating strong performance in classification with balanced precision and recall. LightGBM and Random Forest also performed competitively, with accuracies of 85.33% and 85.25%, respectively. Notably, Random Forest had the highest recall (96.97%) but lower precision (80%). SVM showed strong recall at 93.94% but had the lowest accuracy (83.61%). The findings underscore the importance of model interpretability, facilitated by SHAP, LIME, and ICE, which enhance understanding of model decisions in healthcare applications, ultimately supporting improved clinical outcomes.

1. Introduction

Heart failure (HF) represents a critical global health challenge, impacting millions of individuals and contributing significantly to morbidity and mortality rates worldwide [1]. As a progressive condition characterized by the heart's inability to pump adequate blood to meet the metabolic demands of the body, heart failure encompasses a diverse array of clinical presentations and underlying etiologies. The complexity of this condition necessitates robust diagnostic and management strategies, as traditional clinical assessments often fall short in accurately predicting patient outcomes [2].
In assessing heart health and the progression of heart failure, factors such as an individual's age, gender, blood pressure, fasting blood glucose, and cholesterol levels are of significant importance [3].
The rising prevalence of heart failure, compounded by an aging population and increasing incidence of risk factors such as hypertension, diabetes, and coronary artery disease, underscores the urgent need for innovative approaches to enhance risk stratification and optimize patient care [4]. In this context, machine learning (ML) has emerged as a powerful tool capable of transforming healthcare practices, particularly in the realm of cardiovascular disease. By harnessing large and complex datasets, machine learning algorithms are adept at identifying intricate patterns and relationships that may be obscured by conventional statistical methods [5].
Recent studies have demonstrated the efficacy of various machine learning methodologies in predicting heart failure onset, progression, and clinical outcomes. These algorithms can analyze numerous variables simultaneously, offering a multifaceted understanding of factors influencing heart failure risk [6]. The use of predictive modeling techniques not only enables healthcare professionals to make informed clinical decisions but also supports the creation of personalized treatment strategies adapted to the specific needs of individual patients.
The predictor variables in this analysis, including Age, Sex, ChestPainType, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina, Oldpeak, and ST_Slope, are consistent with the clinical risk factors for heart failure as outlined in the American Heart Association guidelines [7].

This study primarily aims to predict heart failure risk using various machine learning algorithms, including Gradient Boosting Machines (GBM), Support Vector Machine (SVM), Random Forest, and Logistic Regression. In addition, it will evaluate the interpretability of SHAP (SHapley Additive exPlanations), ICE (Individual Conditional Expectation), and LIME (Local Interpretable Model-agnostic Explanations) to clarify the influence of different predictors on patient outcomes. By addressing the current literature's reliance on similar visualization methods, this research introduces a more diverse range of graphical techniques. Through these advanced, interpretable approaches, the study seeks to advance the role of machine learning in clinical practice, ultimately aiming to improve heart failure management and support patient quality of life while reducing healthcare burdens.

2. Material and Methods

2.1 Data Source

The heart failure dataset used in this study was obtained from the publicly available Kaggle database, containing the medical records of 918 patients. Out of these patients, 410 were diagnosed without heart failure, while 508 were diagnosed with heart failure.

2.2 Predictor Variable

Age: Refers to the patient's age in years, a crucial demographic factor influencing health outcomes.
Sex: Male (M) or Female (F).
ChestPainType: Indicates the type of chest pain experienced, classified into Typical Angina (TA), Atypical Angina (ATA), Non-Anginal Pain (NAP), and Asymptomatic (ASY), relevant for diagnosing heart conditions.
RestingBP: Measures resting blood pressure in mm Hg, a key indicator of cardiovascular health.
Cholesterol: Reflects serum cholesterol levels in mg/dl.
FastingBS: Denotes fasting blood sugar levels, where a value of 1 indicates levels greater than 120 mg/dl and 0 indicates otherwise, important for evaluating metabolic health.
RestingECG: Describes the results of a resting electrocardiogram, categorized as Normal, ST (indicating ST-T wave abnormalities), or LVH (suggestive of left ventricular hypertrophy), essential for cardiac assessment.
MaxHR: Represents the maximum heart rate achieved during exercise, a numeric value between 60 and 202, relevant for evaluating cardiac function.
ExerciseAngina: Indicates the presence of exercise-induced angina, marked as Yes (Y) or No (N), crucial for understanding angina-related symptoms.
Oldpeak: Measures the ST segment depression during exercise, represented as a numeric value, important for assessing ischemic changes.
ST_Slope: Describes the slope of the peak exercise ST segment, categorized as Up, Flat, or Down, providing insights into cardiac stress response.
Heart Failure: The output class indicating the presence of heart failure, where 1 signifies heart failure and 0 signifies normal heart function.

2.3 Data Splitting

The dataset was divided into training and test sets with a ratio of 70% to 30%.

2.4 Model Development

In this study, four different machine learning models— Gradient Boosting Machines (GBM), Support Vector Machine (SVM), Random Forest, and Binary Logistic Regression—were applied to analyze the heart attacks data. These algorithms were developed using Python version 3.10.12 to ensure compatibility with the latest libraries and features.
For hyperparameter tuning, the Grid Search technique was utilized. The primary objective of this approach was to improve the performance of each model by determining the optimal combinations of hyperparameters. To achieve this, a k-fold cross-validation method with k=5 was implemented to identify the hyperparameters that would maximize model efficacy.

LightGBM

LightGBM is a gradient boosting algorithm developed by Microsoft, optimized for large datasets and high-dimensional problems. Its histogram-based learning approach significantly reduces training time while optimizing memory usage. By employing a leaf-wise growth strategy, LightGBM enhances the model's generalization performance, leading to better outcomes. Additionally, the Gradient-based One-Side Sampling technique retains samples with larger and more effective gradients during training, thereby improving model performance [8,9].

SVM

Support Vector Machine (SVM) is a widely utilized supervised learning method for classification, introduced by Vapnik et al. in the mid-1990s. It identifies the hyperplane that maximizes the margin between classes, enhancing generalization and reducing overfitting [10]. SVM can transform non-linearly separable data into higher dimensions, allowing for the classification of complex datasets [11]. Its robust performance in high-dimensional spaces makes SVM a powerful tool for various classification challenges.

Random Forest

Random Forest is a supervised learning algorithm and a prominent ensemble method developed by Leo Breiman [12]. It combines multiple decision trees to create a robust classifier, enhancing generalization and minimizing overfitting. Each tree is built using a random subset of training data, and a random selection of features is used at each node to determine splits [13]. This approach increases model diversity, leading to more stable and accurate predictions. By aggregating the outputs of numerous trees, Random Forest reduces variance and enhances overall performance.

Binary Logistic Regression

Binary logistic regression is a statistical model utilized for binary classification tasks, where the dependent variable is categorized into two distinct classes. This model operates on the premise that independent variables exert an influence on the outcome and employs a logistic function to evaluate their effects.

L2 regularization is incorporated to control model complexity by minimizing the sum of the squares of the regression coefficients, thereby mitigating the risk of overfitting. This technique enhances the model's generalizability, resulting in more reliable predictions [14].

2.5 Performance metrics comparison of machine learning algorithms

Four machine learning algorithms were employed to assess and compare model performance based on accuracy, precision, recall, F1 scores, and the area under the curve (AUC) metrics.

2.6 SHAP, LIME, ICE and Interpretability of Machine Learning Models

SHAP: A Game-Theoretic Approach to Feature Attribution

SHAP is grounded in cooperative game theory, employing Shapley values to quantitatively assess the contribution of each feature to model predictions. This method ensures a fair distribution of each feature's influence, allowing for a nuanced understanding of how individual variables drive the model's decision-making process. By providing insights into which features are most significant, SHAP enables researchers to gain a deeper understanding of the underlying factors influencing predictions, particularly in complex models. The SHAP values are often represented in color-coded plots, where the color intensity typically reflects the direction and magnitude of each feature's impact. For instance, red may indicate a positive contribution to the prediction (increasing the likelihood of a heart attack), while blue could signify a negative impact (decreasing that likelihood). This visual representation not only aids in comprehension but also facilitates quicker identification of critical features, enhancing transparency and fostering trust in model outputs [15].

LIME: Local Interpretability for Complex Models

LIME offers an alternative approach by generating local, interpretable models—often linear approximations—that mimic the behavior of more complex models around specific predictions. Through modifications of input data, LIME examines how changes affect model predictions, thereby identifying the most influential features for particular outcomes. The results of LIME analyses can also be depicted using color-coding, where certain colors represent the degree to which features sway the prediction in either direction. This localized interpretability is crucial for understanding the intricacies of complex models, making LIME an essential tool for practitioners seeking to unravel the decision-making processes underlying individual predictions [16].

ICE: Visualizing Individual Feature Impacts

ICE complements SHAP and LIME by focusing on the relationship between individual features and model predictions. By visualizing how alterations in a specific feature's value influence predictions, ICE facilitates a detailed analysis of individual observations. ICE plots often employ colors to indicate different individuals or groups, making it easier to discern trends and interactions. For example, a gradient color scheme might show how different patient groups respond to changes in a particular feature, enhancing the comprehension of how variations in specific features affect overall model performance [17].

3. Results and Discussions

The classification metrics outlined in the table provide a comprehensive comparison of the LightGBM, SVM, Random Forest, and Logistic Regression models applied to the heart failure dataset. (Table 1).

The performance metrics for predicting heart failure using various algorithms reveal valuable insights into their effectiveness. Among the models assessed, Logistic Regression achieved the highest accuracy of 86.89%, indicating its strong overall performance in correctly classifying instances. This model also exhibited a commendable balance between precision and recall, with a precision of 85.71% and a recall of 90.91%. Such results suggest that while Logistic

Regression is effective in identifying positive cases, it also maintains a low rate of false positives.

LightGBM and Random Forest showed competitive results, with accuracies of 85.33% and 85.25%, respectively. LightGBM demonstrated a precision of 87.74% and a recall of 86.92%, making it a robust choice for minimizing false negatives, although slightly less effective than Logistic Regression in overall accuracy. Meanwhile, Random Forest exhibited the highest recall of 96.97%, highlighting its capability in identifying true positive cases, though its precision of 80% indicates a higher occurrence of false positives compared to LightGBM. Both models also yielded strong AUC scores, with Random Forest achieving an AUC of 0.9372, indicating excellent discriminative ability.

Support Vector Machine (SVM) achieved an accuracy of 83.61% but excelled in recall with a score of 93.94%, highlighting its strength in capturing true cases of heart failure. However, it recorded a lower AUC of 0.9026 compared to Random Forest, which suggests that while SVM is effective at identifying true positives, its overall discriminative performance may be less robust. The F1 scores for all models were relatively close, with Logistic Regression leading at 0.8824, closely followed by Random Forest at 0.8767.

Table 1: *Evaluation Metrics for Machine Learning Algorithms*

	SVM	Light GBM	Random Forest	Logistic Regression
Accuracy	0.8361	0.8533	0.8525	0.8689
Recall	0.9394	0.8692	0.9697	0.9091
Precision	0.7949	0.8774	0.8000	0.8571
F1	0.8611	0.8732	0.8767	0.8824
AUC	0.9026	0.8774	0.9372	0.9102

This SHAP summary plot shows Figure 1 the impact of various features on the heart failure prediction model's output. Each point represents a SHAP value for a feature of an individual instance. Features are listed vertically, with the most influential at the top.

Chest Pain Type: The "TA" chest pain type has a positive impact on predicting heart failure (SHAP values > 0), while the "ATA" type contributes negatively to the prediction. This suggests that patients with TA-type chest pain are more likely to be at risk, while those with ATA-type pain may be less likely.

Resting ECG and Resting Blood Pressure: Higher values for the "RestingECG_ST" and lower blood pressure tend to contribute negatively (left side) to the model's prediction. These features are significant, but their SHAP values are centered closer to zero, indicating moderate impact.

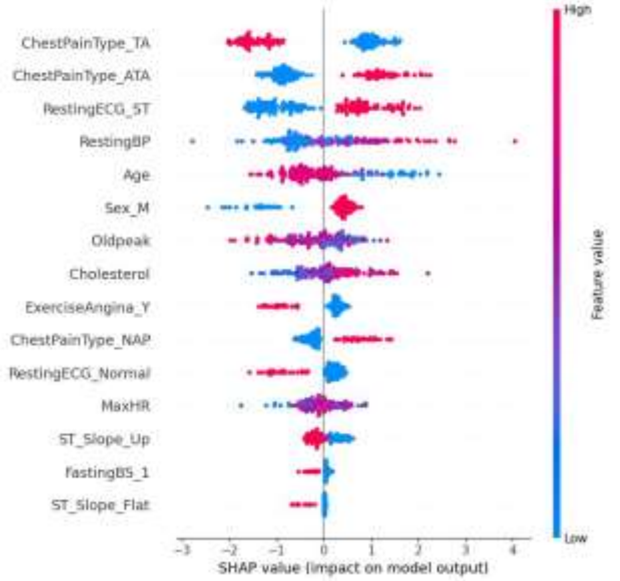

Figure 1. *SHAP-Summary Plot*

The SHAP decision plot illustrates the cumulative effect of each feature on model predictions, showing how individual contributions combine to reach the final output. This visualization provides a clear, step-by-step breakdown of feature impacts, enhancing interpretability, as shown in Figure 2.

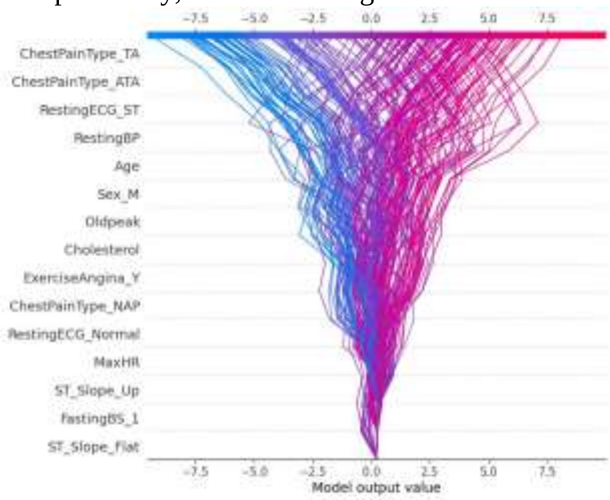

Figure 2. *SHAP-Decision Plot*

The SHAP beeswarm plot in Figure 3 provides an interpretable analysis of the features used by the machine learning model for predicting heart failure. Each feature's SHAP values demonstrate its impact on the model's output, indicating how the feature influences the likelihood of heart failure. Positive SHAP values suggest that a feature increases the risk of heart failure, while negative values suggest a protective effect. The color gradient indicates the feature values, with red representing high values and blue representing low values.

Several key observations can be derived from the SHAP plot:

Chest Pain Type: "ChestPainType_TA" (typical angina) has a strong positive SHAP value for high values, indicating that its presence significantly increases the risk of heart failure. In contrast, "ChestPainType_ATA" (asymptomatic) is associated with lower risk, with negative SHAP values suggesting a protective effect.
Resting ECG and Resting Blood Pressure: Features such as "RestingECG_ST" (ST-T wave abnormality in resting ECG) and "RestingBP" (resting blood pressure) play critical roles in risk assessment. Higher values in "RestingBP" are associated with an increased risk of heart failure, while "RestingECG_ST" high values contribute to a decreased risk.
Age: Age is a significant risk factor, with higher values correlating positively with heart failure risk, as indicated by positive SHAP values for older ages.
Oldpeak: The "Oldpeak" feature, representing ST depression induced by exercise, is another important predictor.
Cholesterol: Cholesterol levels are similarly influential, with high values increasing the risk of heart failure. The importance bar graph illustrating the contribution of each feature is presented in Figure 4. The SHAP force plot visually represents the contribution of each feature toward a specific prediction, highlighting positive and negative influences on the output (Figure 5).

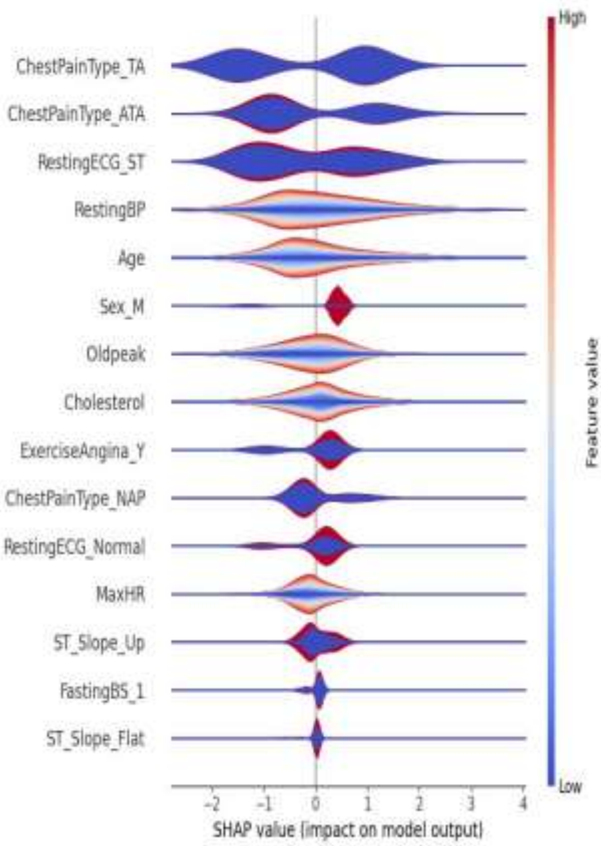

Figure 3. *SHAP Beeswarm Plot*

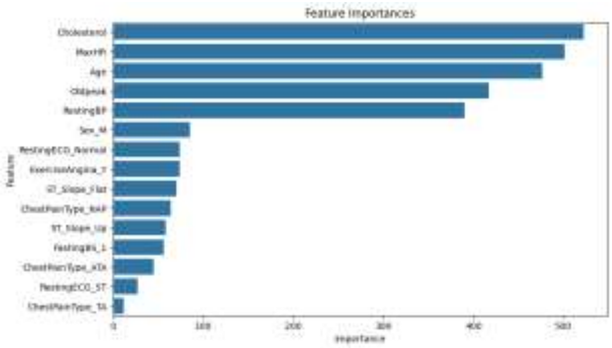

Figure 4. SHAP-*Importance graph*

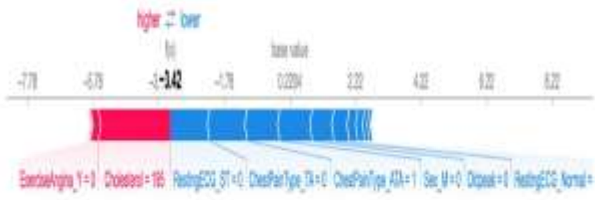

Figure 5. *SHAP Force Plot*

The visualization in Figure 6 displays the model's prediction for heart failure probability, along with the feature contributions that influenced this prediction. The bar chart on the left shows the predicted probabilities, with a high likelihood of heart failure (0.98) compared to no heart failure (0.02).
The right details each feature's contribution to the final prediction. Features with positive contributions (orange) increase the probability of heart failure, while features with negative contributions (blue) decrease it. Notable factors include:
ST_Slope_Flat and ChestPainType_ATA: Both strongly contribute to increasing the heart failure probability, suggesting that a flat ST slope and certain types of chest pain are significant risk indicators.
RestingBP and RestingECG_Normal: These features also contribute to the increased likelihood of heart failure.
Other Features: Features like ST_Slope_Up and ExerciseAngina_Y contribute negatively, slightly reducing the predicted probability of heart failure.

Figure 6. *Lime graph*

The ICE plot in Figure 7 illustrates the relationship between cholesterol levels and the model's predicted probability of heart failure. The blue line represents the model's prediction as cholesterol levels vary, while red points show the distribution of original data points. This plot, demonstrated here for cholesterol, can be similarly generated for all other numerical features to assess their impact on model predictions.

High Cholesterol Levels: For cholesterol levels above 300, the model's predicted probability stabilizes, indicating no further increase in heart failure risk.

Intermediate Cholesterol Levels (100-300): In this range, predictions fluctuate, with a sharp drop around 200. This suggests a non-linear relationship with heart failure risk.

Low Cholesterol Levels: Surprisingly, very low cholesterol levels (below 100) are associated with higher predicted probabilities, possibly due to limited data or interactions with other features.

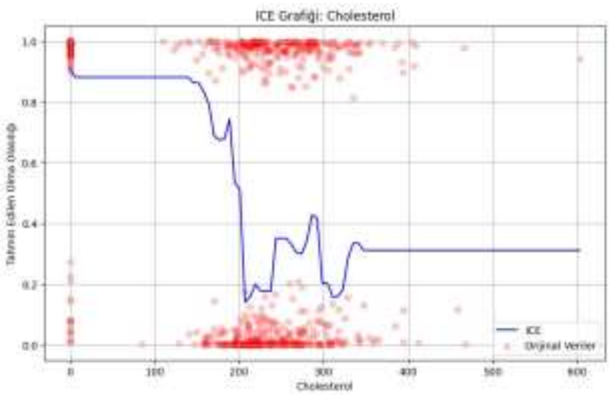

***Figure 7**: Sample Partial Dependence Ice Graph*

3.1 Discussion

In this study, the classification metrics reveal nuanced performance differences across the machine learning models applied to heart failure prediction. Logistic Regression, with the highest accuracy, demonstrates a well-balanced approach to precision and recall, effectively minimizing both false positives and false negatives. LightGBM and Random Forest also achieved competitive results, with Random Forest excelling in recall, indicating its strength in identifying true positive cases. SVM showed strong recall performance as well, though its lower AUC suggests limitations in overall discriminative ability. These findings underscore the importance of selecting algorithms based on the specific requirements of heart failure prediction, such as the need to balance precision with recall or prioritize high sensitivity to positive cases.

For the benefits of the graphs; Beeswarm Plot analysis highlights the features that most strongly influence the model's predictions of heart failure risk, providing insights into the model's decision-making process. Key features such as age, blood pressure, and specific chest pain types significantly affect the risk assessment, with certain feature values (e.g., high blood pressure or specific types of chest pain) being associated with increased risk. This interpretability analysis thus enhances our understanding of the model's predictive behavior and aids in identifying critical factors in heart failure risk assessment.

This ICE plot highlights the complex impact of cholesterol on heart failure predictions, particularly in intermediate ranges. The variability observed across cholesterol levels suggests that cholesterol may interact with other features in determining heart failure risk. This individualized interpretation aids in understanding how the model's predictions vary as cholesterol levels change for a particular instance.

This LIME breakdown provides a clear view of which features most strongly influenced the model's decision, allowing for better interpretability of the heart failure risk prediction.

The SHAP plot visually identifies the importance of each feature, showing how high or low values of these features influence the model's output towards or away from a prediction of heart failure.

El-Sofany and colleagues explored machine learning algorithms for predicting heart failure, achieving notable results. They evaluated ten algorithms, including XGBoost, SVM, and random forests. The XGBoost algorithm demonstrated the best performance, with an accuracy of 97.57%, sensitivity of 96.61%, specificity of 90.48%, precision of 95.00%, F1 score of 92.68%, and AUC of 98% when applied to a combined dataset using the SF-2 feature subset [18]. In our study, we also found that the Random Forest algorithm yielded the best results, underscoring the effectiveness of these machine learning approaches in enhancing early detection of heart failure.

Ahmed and colleagues investigated the role of explainable artificial intelligence (XAI) in enhancing transparency in machine learning models for diabetes prediction. They utilized a logistic regression architecture trained on 253,680 survey responses from diabetes patients. Employing model-agnostic techniques such as LIME and SHAP, the study generated local and global explanations for predictions made by both the logistic regression and Random Forest models on validation and test sets. Their findings revealed a high accuracy of 86% on the test set, highlighting the potential of integrating machine learning with XAI to improve diabetes prediction, diagnosis, and treatment. The comparative analysis of LIME and SHAP underscored their respective strengths and

weaknesses in interpretation, while also addressing future applications and challenges in the field [19]. Additionally, our study includes a detailed explanation of SHAP graph types, LIME, and internal graphics with illustrative examples.
Dave and colleagues explore the role of Explainable Artificial Intelligence (XAI) in enhancing the reliability of AI applications in healthcare. They address concerns related to transparency and model bias, highlighting various interpretability techniques that can improve the understandability of AI systems. Using examples from heart failure datasets, the study emphasizes the importance of XAI in fostering trust in medical diagnostic processes, ultimately supporting broader adoption of AI in healthcare settings and also other publications [20-27].

4. Conclusions

In conclusion, SHAP, LIME, and ICE collectively enhance the interpretability of machine learning models, each offering distinct benefits. SHAP provides comprehensive explanations, LIME delivers localized insights, and ICE visualizes individual feature impacts. This integration is crucial in healthcare applications, where understanding model decision-making is vital for reliable outcomes. By utilizing these tools, practitioners can improve the trustworthiness of machine learning, leading to better clinical decisions and patient care.
The performance metrics from the heart failure dataset highlight the importance of interpretability in model selection. Logistic Regression's highest accuracy of 86.89% and balanced precision and recall demonstrate its effectiveness, while LightGBM and Random Forest also offer strong results. Random Forest's high recall and AUC further emphasize the need for interpretability tools to understand model strengths. This holistic approach is essential for fostering trust in machine learning applications in healthcare.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** I would like to extend my deepest gratitude to Ahmet Öznacar for his invaluable contributions, which were instrumental in the successful completion of this study.

- **Author contributions:** Z.T.S. wrote the introduction section and contributed to the results and discussion. T.Ö. was responsible for writing the methods section, as well as the results section and the discussion.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The dataset is available at https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

References

[1] Ziaeian, B., & Fonarow, G. C. (2016). Epidemiology and aetiology of heart failure. *Nature Reviews Cardiology*, 13(6), 368-378.

[2] Yancy, C. W., et al. (2013). 2013 ACCF/AHA Guideline for the Management of Heart Failure. *Journal of the American College of Cardiology*, 62(16), e147-e239.

[3] Meijers, W. C., & de Boer, R. A. (2019). Common risk factors for heart failure and cancer. *Cardiovascular research*, 115(5), 844-853.

[4] Heidenreich, P. A., et al. (2013). Forecasting the impact of heart failure in the United States: A policy statement from the American Heart Association. *Circulation: Heart Failure*, 6(3), 606-619.

[5] Chaudhary, K., et al. (2020). Machine learning in heart failure: A systematic review. *Journal of Cardiac Failure*, 26(7), 582-590.

[6] Lloyd-Jones, D. M., et al. (2010). Heart disease and stroke statistics—2010 update: A report from the American Heart Association. *Circulation*, 121(7), e46-e215.

[7] Writing Committee Members; (2022). ACC/AHA Joint Committee Members. AHA/ACC/HFSA Guideline for the Management of Heart Failure. *J Card Fail.* 2022 May;28(5):e1-e167. doi: 10.1016/j.cardfail.2022.02.010.

[8] Chen, J., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

[9] Ke, G., Meng, Q., Bai, F., Cheng, J., & Li, Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.

[10] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018

[11] Schölkopf, B., Smola, A. J., & Williamson, R. C. (1999). New support vector algorithms. *Neural Computation*, 12(5), 1207-1245.

[12] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

[13] Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. R News, 2(3), 18-22.

[14] Smith, J. (2020). Understanding L2 Regularization in Logistic Regression. *Journal of Statistical Modeling*, 15(2), 45-59.

[15] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st *International Conference on Neural Information Processing Systems (NIPS* 2017) (pp. 4765-4774). Curran Associates Inc.

[16] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144).

[17] Goldstein, A., Gentile, S., Ghosh, S., & Lee, S. I. (2015). "Peeking Inside the Black Box: A Survey of Explainable AI". *In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2015)*.

[18] Ahmed, S., Kaiser, M. S., Hossain, M. S., & Andersson, K. (2024). A comparative analysis of LIME and SHAP interpreters with explainable ML-based diabetes predictions. *IEEE Access*. Advance online publication. https://doi.org/10.1109/ACCESS.2024.3422319.

[19] El-Sofany, H., Bouallegue, B., & Abd El-Latif, Y. M. (2024). A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. *Scientific Reports, 14*, Article 74656. https://doi.org/10.1038/s41598-024-74656-2.

[20] Dave, D., Naik, H., Singhal, S., & Patel, P. (2020). *Explainable AI meets Healthcare: A Study on Heart Disease Dataset*. arXiv:2011.03195 [cs.LG]. https://doi.org/10.48550/arXiv.2011.03195

[21] Ponugoti Kalpana, L. Smitha, Dasari Madhavi, Shaik Abdul Nabi, G. Kalpana, & Kodati , S. (2024). A Smart Irrigation System Using the IoT and Advanced Machine Learning Model: A Systematic Literature Review. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1158-1168. https://doi.org/10.22399/ijcesen.526

[22] Naresh Babu KOSURI, & Suneetha MANNE. (2024). Revolutionizing Facial Recognition: A Dolphin Glowworm Hybrid Approach for Masked and Unmasked Scenarios. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1015-1031. https://doi.org/10.22399/ijcesen.560

[23] LAVUDIYA, N. S., & C.V.P.R Prasad. (2024). Enhancing Ophthalmological Diagnoses: An Adaptive Ensemble Learning Approach Using Fundus and OCT Imaging. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1541-1551. https://doi.org/10.22399/ijcesen.678

[24] P. Rathika, S. Yamunadevi, P. Ponni, V. Parthipan, & P. Anju. (2024). Developing an AI-Powered Interactive Virtual Tutor for Enhanced Learning Experiences. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1594-1600. https://doi.org/10.22399/ijcesen.782

[25] J. Prakash, R. Swathiramya, G. Balambigai, R. Menaha, & J.S. Abhirami. (2024). AI-Driven Real-Time Feedback System for Enhanced Student Support: Leveraging Sentiment Analysis and Machine Learning Algorithms. *International Journal of Computational and Experimental Science and Engineering,* 10(4);1567-1574. https://doi.org/10.22399/ijcesen.780

[26] S. Leelavathy, S. Balakrishnan, M. Manikandan, J. Palanimeera, K. Mohana Prabha, & R. Vidhya. (2024). Deep Learning Algorithm Design for Discovery and Dysfunction of Landmines. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1556-1566. https://doi.org/10.22399/ijcesen.686

[27] S. Esakkiammal, & K. Kasturi. (2024). Advancing Educational Outcomes with Artificial Intelligence: Challenges, Opportunities, And Future Directions. *International Journal of Computational and Experimental Science and Engineering*, 10(4);1749-1756. https://doi.org/10.22399/ijcesen.799